DOCUMENT RESUME

ED 138 651 TM 006 316

AUTHOR Hamm, Debra W.
TITLE Estimates of Item Characteristics for the Dichotomous and Multiplicative Binomial Rasch Models.
PUB DATE [Apr 77]
NOTE 13p.; Paper presented at the Annual Meeting of the American Educational Research Association (61st, New York, New York, April 4-8, 1977)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Affective Tests; Computer Programs; Goodness of Fit; *Item Analysis; *Mathematical Models
IDENTIFIERS *Rasch Model

ABSTRACT

The results of data analysis based on the multiplicative binomial model (MLTBIN) are compared with dichotomous data analyzed by using the simple logistic model (SLM). Specifically, the paper focuses on the issues of fit to these Rasch Models and estimates of item difficulty. The data indicate that the SLM and MLTBIN are not comparable in the estimates of item fit. Consequently, if Likert scale item analysis is conducted, it should be done using the MLTBIN since the logic and assumptions underlying the items and the model are congruent. Item difficulty estimates using items dichotomized at the midpoint are highly correlated with MLTBIN difficulty estimates. Item difficulties based on dichotomization at the median were found to be erroneous and should be avoided. (RC)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

AERA 77

ED138651

Estimates of Item Characteristics

For the Dichotomous and Multiplicative Binomial Rasch Models

Debra W. Hamm
University of South Carolina

A paper presented to the annual Convention
of the American Educational Research Association
New York, New York, April 1977

TM006 316

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Introduction

Before computer algorithms based on the multiplicative binomial model became available for the analysis of Likert-type scales many researchers simply dichotomized responses to Likert scale items and applied the simple one parameter logistic model. This process has been very commonly used. Thus, it seems important to investigate the comparability of the dichotomous and polychotomous approaches. The purpose of this paper is to compare the results of data analysis based on the multiplicative binomial model (MLTBIN) with dochotomized data analyzed by using the simple logistic model (SLM). Specifically, it will focus on the issues of fit to the model and estimates of item difficulty.

Procedures and Data

Three different psychometric procedures were applied to five affective instruments in this study. First, each data set was analyzed using BICAL (Wright and Mead, 1976), a FORTRAN program for analyzing data according to the multiplicative binomial model of George Rasch (Rasch, 1961).

In the second approach items were dichotomized at their medians and then analyzed using the simple logistic model (CALFIT). For example, if the proportion of individual's selecting each response to an item were as follows,

1	.10	
2	.15	0
3	.20	
---	---	---
4	.55	1

the item would be dichotomized 1, 2, 3 = 0 and 4 = 1, because this split most closely approximates a division at the 50th percentile for the item.

The third technique of analysis involves items dichotomized at their midpoints. Since this procedure can be applied only to items with an even number of scale points, three of the data sets alone could be analyzed using this technique. For the four-point scales the items were dichotomized between responses two and three. For the six-point scale the items were dichotomized between responses three and four. The items were then analyzed using the simple logistic model.

Three major questions were asked of these analyses.

1. Do all three procedures identify the same items as fitting or misfitting the Rasch model?

2. Are the estimates of the fit of the items to the model (fit mean squares) similar?

3. Are the item difficulty estimates which result from the three procedures similar?

The data used in this study consist of five affective measures. All are composed of Likert scale type items. 1) The first is an academic self concept measure (ACSC) which was administered to 185 sixth grade subjects. The instrument consists of 15 items, each a four-point Likert scale.

2) The second data set is a measure of academic self-concept in language arts. The language arts self-concept instrument (LASC) contains 12 items, each with four scale points.

3) The third data set was collected on a sample of 205 graduate students enrolled in a televised educational research course at a southern university. The instrument was designed to measure attitudes toward educational research (AEDR) and consists of 22 items, each with 5 scale points.

4) A student course evaluation questionnaire (SCE) administered to 428 graduate and undergraduate students constitutes the fourth data set. The student course evaluation questionnaire has 12 items, each with five scale points.

5) The last data set is a subset of items from a study of discrimination and prejudice. In this study, 17 items dealing with responses to women's roles (RWR) were administered as part of a larger instrument. The responses to the 17 item subscale were available for 193 subjects.

Results

Item Fit

The three psychometric procedures are very dissimilar in the items which they identify as misfitting the model. Table one shows the items which each procedure identified as misfitting the model. (This was operationally defined as a fit mean square greater than three.) These analyses indicate that item fit in either of the dichotomized procedures does not assure fit to the multiplicative model nor does fit to the multiplicative model necessarily imply fit to the simple logistic model.

Table I

Misfitting Items
Identified by the MLTBIN
and Simple Logistic Analyses
(Fit Mean Squares are in Parentheses)

	MLTBIN	MEDIAN	MIDPOINT
ACSC	All Items Fit	6(3.02) 8(4.52) 12(4.20) 15(6.49)	1(6.53) 4(3.27) 12(3.77) 15(6.03)
LASC	5(8.21)	5(22.42)	4(4.10) 5(36.92) 11(3.44)
SCE	6(5.43) 11(3.18)	4(4.85)	No midpoint— 5 point scale
AEDR	2(12.07) 4(5.77) 5(4.07) 9(3.37) 12(3.20)	2(9.50) 4(15.92) 5(3.79) 6(4.57)	No midpoint— 5point scale
	2(4.09) 3(5.61) 4(3.06) 5(6.81) 6(3.09) 7(7.24) 8(6.41) 9(3.12) 12(3.29)	7(3.71) 9(3.41)	7(3.28) 11(3.63) 12(3.25) 13(3.36)

Since a multiplicative binomial model seems more congruous with Likert scaling than a psychometric method based on a dichotomy, it might be expected that more items would fit the binomial coefficient model than the simple logistic model. An examination of Table One shows that this is not the case.

For the SCE, AEDR and RWR, more items misfit when analyzed using the multiplicative model than in either of the dichotomized cases. On the other hand, all items fit in the MLTBIN analysis of ACSC, but four items misfit in each analysis of dichotomized items.

The correlations between the fit mean squares for the items as calculated in the three routines were also considered in order to investigate the similarity of the magnitudes of the misfit statistics generated by each procedure. The results provide little evidence to support the comparability of the three methods under consideration. (See Table 2). Only for one instrument, the RWR, are the correlations genuinely significant. For this measure there is a correlation of .519 between the fit mean squares from the MLTBIN analysis (FMSMLTBIN) and the fit mean squares from the analysis where items were dichotomized at the median (MSFDMED). A correlation of .482 was found between the fit mean squares when items were dichotomized at the midpoint (MSFDMID) and MSFDMED. Both of these correlations are significant at the .05 level.

For the language arts self-concept there appears to be a very strong relationship between the fit estimates based on the three analyses. The correlations, however, are deceptively large and are highly inflated by an outlier, item 5. When this item is eliminated the correlations are reduced to -.16 for MSFMLTBIN-MSFDMED, -.26 for MSFMLTBIN-MSFDMID and .09 for MSFDMID-MSFDMED. A similar explanation applies for the AEDR. When items 2 and 4, items with extreme misfit, are eliminated the correlation drops to .46.

Item Difficulties

The analysis of the item difficulty estimates shows that item difficulties

Table Two

Correlations of Mean Square Fit Statistics

	Correlation of MSFMLTBIN and MSFMEDIAN	Correlation of MSFMLTBIN and MSFMIDPOINT	Correlation of MSFMIDPOINT And MSFMEDIAN
ACSC	.368	.345	.398
LASC	.971***	.960***	.986***
SCE	.172	-	-
AEDR	.717**	-	-
RWR	.519*	.281	.482*

*p<.05

**p<.01

***p<.001

Table Three

Correlations of MLTBIN Difficulties and Difficulties
When Items are Dichotomized at the Midpoint

	r	
ACSC	.950	****
LASC	.965	****
RWR	.938	****

**** p .0001

Table Four

Correlations of MLTBIN Difficulties and Difficulties
When Items are Dichotomized at the Median

ACSC	.099	
LASC	.009	
AEDR	.297	*
SCEV	.630	***
RWR	.889	***

* p<.05

*** p<.001

based on dichotomizing at the midpoint are highly correlated with those obtained using MLTBN. (The correlations are shown in Table 3.) For the three tests with an even number of alternatives and hence midpoints on their scales, the correlations were very high and significant in all cases.

The correlation between MLTBIN difficulty estimates and estimates based on dichotomizing at the median were generally much lower than those between MLTBIN difficulties and midpoint dichotomized difficulties. (See Table 4.)

The major problem with difficulty estimates based on the median dichotomy is that it makes all item difficulty estimates similar. Theoretically all should have a classical item difficulty of .50 and a difficulty of 0.0 in the logit metric. An extremely negatively skewed item and an extremely positively skewed item, if dichotomized precisely at their medians, would be estimated to have the same difficulty. This of course presents a faulty picture of the data. A correct analysis would find the negatively skewed item to be "easy" and the positively skewed item to be "hard".

The problem encountered when using items dichotomized at the median to estimate item difficulties can be clearly illustrated by an example. Consider items 2 and 10 on the academic self-concept instrument. The number of individuals selecting each response for these 2 items is shown in Table 5. In this table the arrows indicate where the items were dichotomized. Table 6 shows the difficulty estimates from each procedure.

Table 5

Number of Subjects Selecting Each Response

Item 2	Alternative No.	1	2	3	4
	Frequency	13	88	74	8
Item 10	Alternative No.	1	2	3	4
	Frequency	101	66	8	6

Table 6

Item Difficulty Estimates

Item	MLTBIN	MIDPOINT	MEDIAN
2	-.537	-1.686	.03
10	.365	1.347	.03

When these items are dichotomized at the median they both have a difficulty of +.03. In the case of item two this estimate probably represents the difficulty of the item fairly accurately. The distribution of the item responses is roughly normal and most people have selected moderate responses. Thus, a fairly average difficulty is reasonable.[1]

Now consider item 10. The distribution for this item is very negatively skewed. (On this scale low scores represent positive affect.) The shape of the item distribution suggests that this is a very "easy" item. One hundred and one people, the majority, selected the most positive response. However, this item was estimated to have the same difficulty, +.03, as item two.

[1]The item difficulties are deviations from the mean of the item difficulties. Consequently, the closer the item difficulties are to zero, the more average their difficulty.

This occurred because when the items were dichotomized at their medians it resulted in exactly the same number of people having the "correct" response on each of the items. For both items two and ten there were 101 people who made the "correct" response and 82 who made the "incorrect" response. If we were to dichotomize all the items at precisely their medians they would each have exactly the same difficulty. For this reason, difficulty estimates based on the median are not useful or accurate. If items must be dichotomized the midpoint is a much better choice than the median if difficulties are to be estimated. This is because dichotomization at the midpoint will provide estimates that reflect the distribution of responses to the items whereas the median dichotomy will obliterate these important differences. Furthermore, the midpoint dichotomy seems more consistent with the multiplicative model's assumption of equal spacing of response alternatives.

Summary and Conclusions

The data indicate that the simple logistic model and multiplicative model are not comparable in the estimates of item fit. Consequently, if Likert scale item analysis is concluded, it should be done using the multiplicative binomial model since the logic and assumptions underlying the items and the model are congruent.

Item difficulty estimates using items dichotomized at the midpoint are highly correlated with MLTBIN difficulty estimates. Item difficulties based on dichotomization at the median were found to be erroneous and should be avoided.

References

Rasch, G.: On general laws and the meaning of measurement in psychology. In Procedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability. IV, Berkeley, University of Californiz Press, 1961, 321-34.

Wright, B. D. and Mead, R.: BICAL: Calibrating Items and Scales with the Rasch Model. Research Memorandum #23, Educational Statistics Laboratory, University of Chicago, 1976.

Wright, B. D. and Mead, R.: CALFIT Sample-Free Item Calibration with a Rasch Measurement Model. Research Memorandum #18, Educational Statistics Laboratory, University of Chicago, 1975.